UNIVERSITÉ DE CAEN
U.F.R. de Sciences

**L3 Informatique**
**Systèmes et réseaux**

DEVOIR 1 — 2010

---

# But

Montrer votre maîtrise des outils de manipulations de données texte sous Unix ainsi que du shell et awk, seuls langages autorisés pour ce devoir. Vous vous limiterez le plus possible au Bourne shell standard. Vous n'utiliserez les facilités supplémentaires du bash que si cela est strictement nécessaire et justifierez leur utilisation dans votre rapport qui sera réalisé *obligatoirement* avec LaTeX. Vous devrez également utiliser le programme `gnuplot` pour dessiner des graphiques.

Une écriture structurée des programmes sera très appréciée.

# Journaux de processeurs gestionnaires de courrier électronique

Vous trouverez sur les machines du réseau dans le fichier `/home/data/l3/systeme/mail-anonyme.log.bz2` un extrait de journal des serveurs de courrier électronique. Il s'agit d'un extrait des journaux collectés au département d'informatique. Ces traces ont été anonymisées, c'est-à-dire que tous les adresses email ou IP ont été remplacées.

Il contient les traces des programmes suivants :
- archived
- bounced
- courierd
- courieresmtp
- courieresmtpd
- courierlocal
- couriertcpd
- esmtpd-ssl
- imapd-ssl
- pop3d
- pop3d-ssl
- spamd
- sqwebmaild
- sympa
- task_manager
- wwsympa

**Question 1** *Écrire un script qui produise cette liste de programmes*

Il s'agit maintenant de produire quelques statistiques d'utilisation en analysant ces traces.

**Question 2** *Calculez le nombre par jour et le nombre total sur la période d'erreurs du protocole* `smtp`*.*

**Question 3** *Calculez le nombre d'utilisateurs différents par heure et par jour des services de* `imap` *et* `pop`*.*

*Calculez le nombre d'octets transférés par chacun de ces protocoles, par heure et par jour.*

**Question 4** *Calculez la durée moyenne d'une connexion au service* `imap` *et* `pop`*.*

**Question 5** *Permette la constitution de rapports de tous les résultats demandés sous forme de tableau et graphique en* HTML *ou en* PDF*. Le* PDF *sera généré en utilisant* LaTeX*. Le graphique sera produit grâce à* `gnuplot`*.*

Enfin, les extraits de journaux mis à votre disposition ont été anonymisés. Chaque adresse mail ou IP a été remplacée par une autre chaîne syntaxiquement correcte (et ici, de même longueur et correspondant à un mot de dictionnaire, c'est-à-dire certainement prononçable). Attention il faut qu'une adresse originale soit remplacée partout où elle apparait dans le journal par la même chaîne de caractères sous peine de ne pas pouvoir faire ultérieurement certaines analyses (par ex. durée de connexion).

**Question 6** *Écrire un script (ou un ensemble de script) qui anonymise les traces. On ne demande pas l'anonymisation des adresses IP.*

## Travail à rendre

Le travail est à faire par groupe de deux étudiants.

Vous devez écrire des programmes en shell ou awk réalisant les calculs et transformations demandés à partir des journaux. Ces calculs devront pouvoir être lancés par un programme acceptant en argument sur sa ligne de commande le nom du journal ainsi que des arguments contrôlant quelle statistiques calculer et sous quelle forme les rendre. Vous réaliserez en shell à l'aide de `dialog` un programme frontal du précédent et d'utilisation plus conviviale.

Vous devez rendre avant le 23/10/2010, 20h délai de rigueur, un dossier réalisé avec LaTeX comprenant :

1. une courte explication générale du fonctionnement de vos scripts ;
2. un manuel d'utilisation à la Unix de vos programmes ;
3. les sources de vos programmes raisonnablement commentés ;
4. des exemples de tests d'exécutions ;
5. joindre également le source LaTeX de votre rapport.

## Modalités de remise du travail

Tous les fichiers utilisés seront regroupés dans une archive compressée en `tar.gz` qui sera déposée sur le système à l'adresse `http://devoirs.info.unicaen.fr/`. Cette archive comprendra toujours, à sa racine, un fichier nommé `noms.txt` contenant les deux noms du binôme étudiant. Seul un des deux étudiants rendra une archive avec les fichiers du devoir ; l'autre rendra une archive contenant seulement le fichier obligatoire `noms.txt`.

Tant que la date limite de remise n'est pas dépassée, vous pouvez remplacer l'archive par une nouvelle. Vous ne pourrez pas la modifier après cette date limite. Une remise après cette date est toutefois possible si elle n'avait pas été précédée d'une autre ; ce retard sera naturellement pénalisé.

Pour vous identifier sur le serveur devoirs, vous utiliserez votre EtuP@ss.